



Diagnostic value of a dynamic artificial intelligence ultrasonic intelligent auxiliary diagnosis system for benign and malignant thyroid nodules in patients with Hashimoto thyroiditis

Bing Wang^{1#}, Zheng Wan^{1#}, Mingbo Zhang², Fengxia Gong¹, Linlin Zhang¹, Yukun Luo², Jing Yao¹, Chen Li¹, Wen Tian¹

¹Senior Department of General Surgery, the First Medical Center of Chinese People's Liberation Army (PLA) General Hospital, Beijing, China;

²Department of Ultrasound, the First Medical Center of Chinese People's Liberation Army (PLA) General Hospital, Beijing, China

Contributions: (I) Conception and design: C Li, W Tian; (II) Administrative support: C Li; (III) Provision of study materials or patients: C Li, W Tian; (IV) Collection and assembly of data: B Wang, Z Wan, M Zhang, F Gong, L Zhang, Y Luo, J Yao; (V) Data analysis and interpretation: B Wang, Z Wan, M Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jing Yao; Chen Li; Wen Tian. Senior Department of General Surgery, the First Medical Center of PLA General Hospital, 28 Fuxing Road, Haidian District, Beijing 100853, China. Email: yaojing301@163.com; lichen_pla301@163.com; tianwen301_cta01@163.com.

Background: A dynamic artificial intelligence (AI) ultrasonic intelligent assistant diagnosis system (dynamic AI) is a joint application of AI technology and medical imaging, which can conduct real-time synchronous dynamic analysis of nodules from multiple sectional views with different angles. This study explored the diagnostic value of dynamic AI for benign and malignant thyroid nodules in patients with Hashimoto thyroiditis (HT) and its significance in guiding surgical treatment strategies.

Methods: Data of 487 patients (154 with and 333 without HT) with 829 thyroid nodules who underwent surgery were collected. Differentiation of benign and malignant nodules was performed using dynamic AI, and diagnostic effects (specificity, sensitivity, negative predictive value, positive predictive value, accuracy, misdiagnosis rate and missed diagnosis rate) was assessed. Differences in diagnostic efficacy were compared among AI, preoperative ultrasound based on the American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS), and fine needle aspiration cytology (FNAC) diagnoses.

Results: The accuracy, specificity and sensitivity of dynamic AI reached 88.06%, 80.19%, and 90.68%, respectively; besides, there was consistency with postoperative pathological consequences ($\kappa=0.690$; $P<0.001$). The diagnostic efficacy of dynamic AI was equivalent between patients with and without HT, and there were no significant differences in sensitivity, specificity, accuracy, positive predictive value, negative predictive value, missed diagnosis rate, and misdiagnosis rate. In patients with HT, dynamic AI had significantly higher specificity and a lower misdiagnosis rate than did preoperative ultrasound based on the ACR TI-RADS ($P<0.05$). Compared with FNAC diagnosis, dynamic AI had a significantly higher sensitivity and a lower missed diagnosis rate ($P<0.05$).

Conclusions: Dynamic AI possessed an elevated diagnostic worth of malignant and benign thyroid nodules in patients with HT, which can provide a new method and valuable information for the diagnosis and development of management strategy of patients.

Keywords: Artificial intelligence (AI); Hashimoto thyroiditis (HT); thyroid nodules; ultrasound examination

Submitted Aug 25, 2022. Accepted for publication Mar 22, 2023. Published online Apr 04, 2023.

doi: 10.21037/qims-22-889

View this article at: <https://dx.doi.org/10.21037/qims-22-889>

Introduction

In recent years, the incidence rate of thyroid nodules has increased. In normal adults, one-third or more thyroid nodules are found in ultrasound examinations. Because 5–8% of thyroid nodules may be cancerous, the main task of initial evaluations of thyroid nodules is to differentiate between benign and malignant nodules in order to find malignant tumors. Ultrasound imaging is the most commonly used method for screening thyroid nodules. With improvements in ultrasound resolution, the incidence rate of thyroid cancer has shown a significant upward trend worldwide (1). Hashimoto thyroiditis (HT) is an autoimmune disease with genetic susceptibility. The incidence rate of HT is 6–10% and is increasing year by year (2-4). The histological characteristics of HT are lymphocyte infiltration and lymphoid follicle destruction. At present, HT is considered to be closely related to the occurrence of thyroid papillary carcinoma (5-7). The typical sonogram of HT shows diffuse changes. The echogenicity of glands is thickened and reduced, and dot-like, sheet-like hypoechoic, or grid-like changes are seen. When nodules are combined, this diffuse change will affect the characteristics of the nodules. Many benign nodules show malignant signs, such as an unclear boundary, irregular shape, and calcification, which increases the difficulty of judging the nature of the nodules and may lead to unnecessary puncture biopsies. One study reported that the accuracy rate for a doctor's diagnosis of thyroid cancer against the background of HT was only 54.28% (8). Although fine needle aspiration cytology (FNAC) is currently an effective means of diagnosing benign and malignant thyroid nodules against the background of HT (9), samples may be affected by HT, and they may thus be unqualified or lack clear significance, resulting in low diagnostic efficacy.

The combination of artificial intelligence (AI) with medical imaging had led to the development of a static AI ultrasound intelligent diagnosis system that not only ensures accuracy, but also improves diagnostic efficacy. This system has been widely used in clinical practice for prenatal examinations (10), cardiac ultrasound examinations (11), and the diagnosis of breast (12) and thyroid (13-15) nodules. However, only a single view of the nodule can be diagnosed by static AI, and the nature of the nodule cannot be judged in real time. One study (16) showed that the sensitivity, specificity and accuracy of computer-assisted diagnosis system in diagnosing thyroid nodules in HT patients were 91.2%, 65.0% and 80.8%, with high sensitivity, but low specificity. There may be over-diagnosis when used alone. In HT

patients, the echogenicity of thyroid parenchyma is always heterogeneous, which make it difficult to differentiate from “real” thyroid nodules. In our previous study (17), we found that dynamic AI had good performance in the diagnosis of thyroid nodules, with a sensitivity of 92.21%, a specificity of 83.20%, an accuracy of 89.97%, and a high consistency with postoperative pathology ($\kappa=0.737$; $P<0.001$). Dynamic AI, which is safe and noninvasive, can be used to perform dynamic analysis of nodules in real time, and improve the efficacy of clinical examinations and diagnosis. Therefore, it is necessary to explore the diagnostic efficacy of dynamic AI for thyroid nodules in HT patients, and provide a new method for the diagnosis of thyroid nodules in clinical HT patients.

In the present study, through a comparison and analysis on dynamic AI diagnostic efficacy with preoperative routine FNAC and ultrasound, the diagnostic value of dynamic AI for benign and malignant nodules in patients with HT and its significance in guiding surgical treatment strategies were evaluated. The following article is presented in accordance with the STARD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-889/rc>).

Methods

Patients

In all, 487 patients with 829 thyroid nodules who underwent surgery in the First Medical Center of the Chinese PLA General Hospital from November 2021 to March 2022 were attained to the study. Age, gender, size, and pathology concerning the thyroid nodules, as well as the results of dynamic AI, preoperative routine ultrasound, and FNAC diagnosis were reviewed retrospectively. To be eligible for inclusion in the study, patients had to meet the following criteria: (I) with thyroid nodules and surgical indications initially diagnosed by physical or auxiliary check-up before surgery, (II) clear postoperative pathological results available, and (III) the complete clinical pathological data and ultrasonic imaging data. Patients with incomplete clinical pathological or ultrasound image data, those with unclear postoperative pathological results, those with FNAC class I (unsatisfactory or nondiagnostic) or class III (atypia of uncertain significance) results, and those with other types of malignant tumors were excluded from the study. *Figure 1* shows the flowchart of patient enrollment.

This study was conducted following the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of PLA General Hospital (No. S2022-

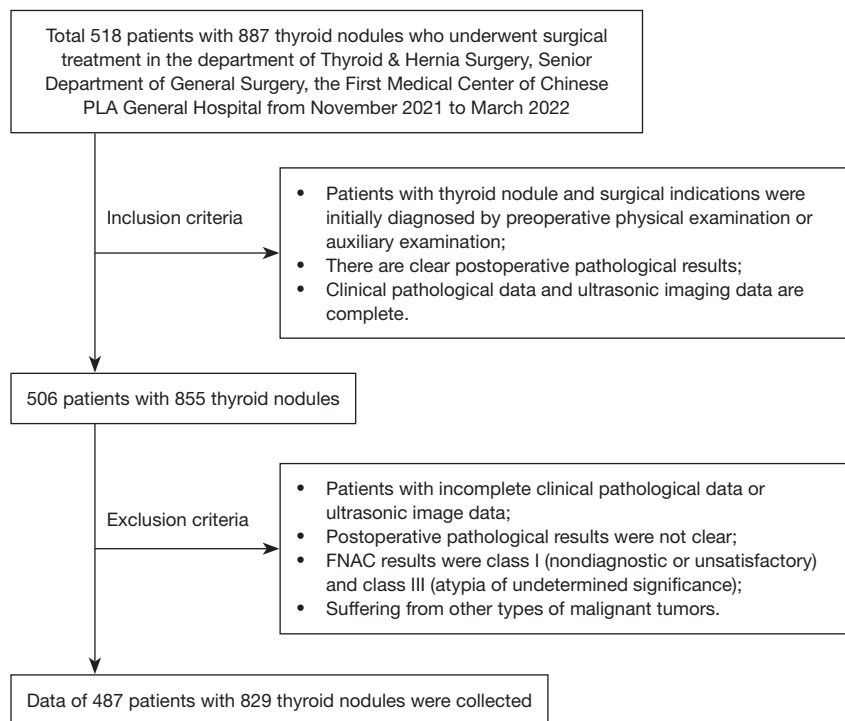
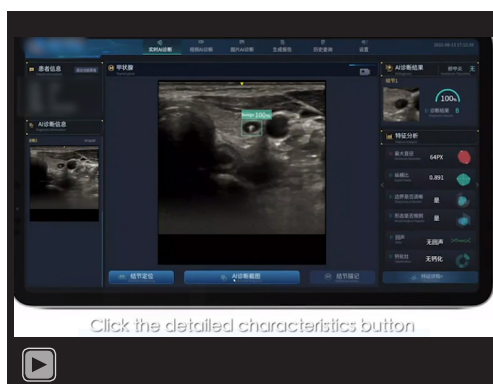


Figure 1 Flow chart showing patient enrollment. FNAC, fine needle aspiration cytology.



Video 1 Operation instructions of dynamic AI.

441-01). Written informed consent was obtained from all participants.

Dynamic AI diagnosis

Dynamic AI diagnoses were obtained using a system of Ian Thyroid Solution 100 (ITS100) (MedAI Technology Co. Ltd, Wuxi, China). This system is a dynamic ultrasound

intelligent assistant real-time diagnostic system comprising a GE LOGIQ e ultrasound device, a probe of GE L4-12T-RS linear array, a host computer, along with an AI auxiliary screen for display (*Video 1*). The whole patients were diagnosed by dynamic AI before operation. In these examinations, patients lay in a supine and position with their neck exposed. According to the actual situation of thyroid nodules in different patients, the depth, focus and the other parameters were properly regulated with a view to procuring the most gratifying image figures. During an examination, the thyroid tissue of the preoperative patient was carefully examined in the following order: from the left to right lobe, between the top and the bottom, and from the inside out. Longitudinal and transverse sections of the thyroid were scanned continuously three times so as to avoid missing marginal lesions and nodules. When a nodule was found, the ITS100 automatically located the nodule real-timely. The deep learning technique is employed in Dynamic AI so as to automatically gain the image characteristics. Based on ultrasound image unique features, it devises an emerging convolutional neural network (CNN) construct for the diagnostic pattern. With the help of the

CNN convolution kernel, the pixels on the input image are sampled to obtain the global image characteristics for the thyroid nodule area in ultrasound image and build an elevated-throughput, multilevel characteristic space. After the diagnostic pattern has performed feature extraction and calculation on the image of the input nodule, the pattern outputs two probability values, representing the possibility that the diagnostic pattern considers the nodule of input to be benign or malignant. As the forecast possibility of the nodule being malignant exceeds or equates to the probability of the nodule being benign, the prediction consequences of the pattern is “malignant” (red “M” sign); Or it’s benign (green “B” sign). The proportion value stands for the possibility of the nodule being malignant or benign. Consequences of final dynamic AI diagnosis are procured by means of integrated dissection combining the consequences of 3 scans. In the present study, the operation is overseen and instructed by one application engineer of AI Corporation in a standardized way. The nature of all nodules was confirmed by pathological results after surgery.

In this study, we analyzed and compared the dynamic AI diagnostic efficacy, preoperative routine FNAC and ultrasound, and evaluated the diagnostic value of dynamic AI in telling benign nodules from malignant nodules in patients with HT. In addition, differentiating between benign and malignant category 4 nodules under the American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) (TR 4 nodules) (18), which have a malignancy risk of 5–20%, is difficult using ultrasound, and this is further complicated in the in the presence of HT. Thus, to better judge nodules as benign or malignant and to provide patients with an individualized treatment plan, we particularly evaluated the use of dynamic AI for TR 4 nodules in patients with HT who had undergone preoperative conventional ultrasound with the aim of developing a new method for the diagnosis of thyroid nodules.

Preoperative routine ultrasound diagnosis

Preoperative routine ultrasound was implemented by means of an EPIQ 7 (Philips Healthcare, Andover, MA, USA) with one L125 linear-array transducer (5–12 MHz). All subjects went through the ultrasound examination prior to surgery. As described above, patients lay supine with their neck exposed. The thyroid gland longitudinal and transverse sections were scanned using an ultrasound probe: firstly, the left lobe; next, the right lobe, between top and bottom, from the inside out. These ultrasound examinations for patients

were finished by the identical doctor. Thyroid nodules were categorized according to ACR TI-RADS (18); categories 1–3 were categorized as benign and categories 4 and 5 were classified as malignant. After patients had been diagnosed by specialized ultrasound doctors, 3 senior ultrasound doctors were invited to review the films. In case of disagreement, the final result was the one that 2 of the 3 doctors agreed on.

Preoperative FNAC diagnosis

The Mylab Twice US mechanism (Esaote SpA, Geneva, Italy) with one L523 linear-array transducer (4–13 MHz) was used for the FNAC examinations. FNAC examinations were conducted in patients with suspected malignant nodules (i.e., ACR TI-RADS grade 4 and above nodules) based on routine ultrasound. For FNAC, patients lay supine with their neck was exposed. A point of the puncture was chosen in the ultrasound guidance, and one 25-G puncture needle was inserted into the nodule center in the ultrasound guidance. Then, the needle was inserted 5 times–10 times in disparate nodule parts. After that needle had been withdrawn, that tissue was used for the cytological smears. Thyroid nodules were classified using the Bethesda reporting system (19) as follows: I, nondiagnostic or unsatisfactory; II, benign; III, atypia of uncertain significance; IV, suspicious for follicular nodule or follicular nodule; V, suspicious for malignant nodule; and VI, malignant nodule. Because FNAC can’t correctly diagnose malignant and benign follicular nodule, for this research, classes II and IV were evaluated as benign; besides, classes V and VI were regarded as malignant. After patients had been diagnosed by specialized doctors, senior pathologists were invited to review the results. In case of disagreement, the final result was the one that 2 of the 3 doctors agreed on.

Postoperative clinical pathological diagnosis

According to the classification of endocrine and neuroendocrine tumors published by the World Health Organization (WHO) (20), the same senior expert in the field of pathology in our hospital read and diagnosed the film.

HT diagnosis

HT was confirmed on the basis of clinical signs, serological examinations, and postoperative pathology, as well as concurrent diffuse lesions appearing on ultrasound.

Statistical analysis

All statistical analyses were performed using SPSS version 23.0 (IBM Corp., Armonk, NY, USA). Based on postoperative pathology, the diagnostic efficacy of dynamic AI, consisting of specificity, sensitivity, negative predictive value (NPV), positive predictive value (PPV), accuracy, misdiagnosis rate and missed diagnosis rate was determined. These outcomes were calculated as follows: sensitivity = number of accurately diagnosed malignant nodules/total number of malignant nodules $\times 100$; specificity = the number of accurately diagnosed benign nodules/the total number of benign nodules $\times 100$; accuracy = the number of accurately diagnosis nodules/the total number of nodules $\times 100$; NPV = the number of accurately diagnosed benign nodules/the number of diagnosed benign nodules $\times 100$; PPV = the number of accurately diagnosed malignant nodules/the number of diagnosed malignant nodules $\times 100$; missed diagnosis rate = the number of misdiagnosed malignant nodules/the total number of malignant nodules $\times 100$; and misdiagnosis rate = the number of misdiagnosed benign nodules/the total number of benign nodules $\times 100$. The conformance of dynamic AI, FNAC, ACR TI-RADS and pathological examinations was analyzed using the kappa test ($\kappa < 0$, no conformance; $\kappa \leq 0.40$, poor conformance; $0.40 < \kappa \leq 0.60$, moderate conformance; $0.60 < \kappa \leq 0.80$, elevated conformance; $\kappa > 0.8$, exceedingly elevated conformance). Frequencies were compared between groups using the χ^2 test. All P-values are two-tailed, and a $P < 0.05$ was considered statistically significant.

Results

In all, 487 patients (132 males, 355 females) were included in this study (Figure 1). Patients ranged in age between 16 years old and 75 years old, with a mean age 43.5 ± 11.6 years. Information was collected for 829 nodules, the diameter of which ranged from 0.2 to 6.5 cm with an interquartile range 0.5 (0.7–1.0 cm) [p50 (p25, p75)]. Of the 487 patients, 154 (243 nodules) were complicated with HT; the remaining 333 patients (586 nodules) were not. All malignant and benign nodules were certified by the postoperative pathology (Figure 2).

Diagnostic efficacy of dynamic AI

Among the 829 nodules in this study, dynamic AI identified 605 malignant and 224 benign nodules, with the diagnosis of 166 benign and 564 malignant nodules corresponding to

the postoperative pathological consequences [$\kappa = 0.690$ (high consistency); $P < 0.001$; Table 1]. The dynamic AI diagnostic efficacy is presented in Table 2.

Stability analysis of dynamic AI in patients with HT

Accuracy of dynamic AI examination by gender

In this study, 154 patients had HT (243 nodules), and 26 of these patients were male (39 nodules). Among males, the dynamic AI diagnosis of 33 of 39 nodules was consistent with postoperative pathological results (accuracy 84.62%). Among the 128 female patients, the dynamic AI diagnosis of 184 of 204 nodules was consistent with postoperative pathological results (accuracy 90.20%). There was not any significant distinction in the dynamic AI accuracy between genders (Table 3).

Accuracy of dynamic AI by age

In the 128 patients with HT aged < 55 years, the dynamic AI diagnosis of 179 of 199 nodules was consistent with the postoperative pathological results (accuracy 89.95%). In the 26 HT patients aged ≥ 55 years, the dynamic AI diagnosis of 38 of 44 nodules was consistent with the postoperative pathological results (accuracy 86.36%). There was no significant difference in the accuracy of dynamic AI among the 2 age groups (Table 3).

Accuracy of dynamic AI by nodule type

Dynamic AI accurately identified 181 of 197 malignant nodules and 36 of 46 benign nodules, with an accuracy of 91.88% and 78.26%, respectively. There was a significant difference in the accuracy of dynamic AI according to the nature of the nodules ($P = 0.007$; Table 3).

Accuracy of dynamic AI by nodule size

The dynamic AI accuracy in identifying nodules that were ≤ 0.5 , > 0.5 –1.0, > 1.0 –2.0, and > 2.0 cm was 85.23%, 90.20%, 97.22%, and 88.24%, respectively. There existed no significant distinction in the dynamic AI accuracy across different-sized nodules (Table 3).

Diagnostic efficacy of dynamic AI in HT and non-HT patients

Of 487 patients, 154 had HT and 333 did not. Of the 243 nodules in those with HT, dynamic AI diagnosed 181 malignant and 36 benign nodules, which highly corresponded to the postoperative pathological consequences [$\kappa = 0.668$ (high consistency); $P < 0.001$; Table 1]. Of the 586 nodules in

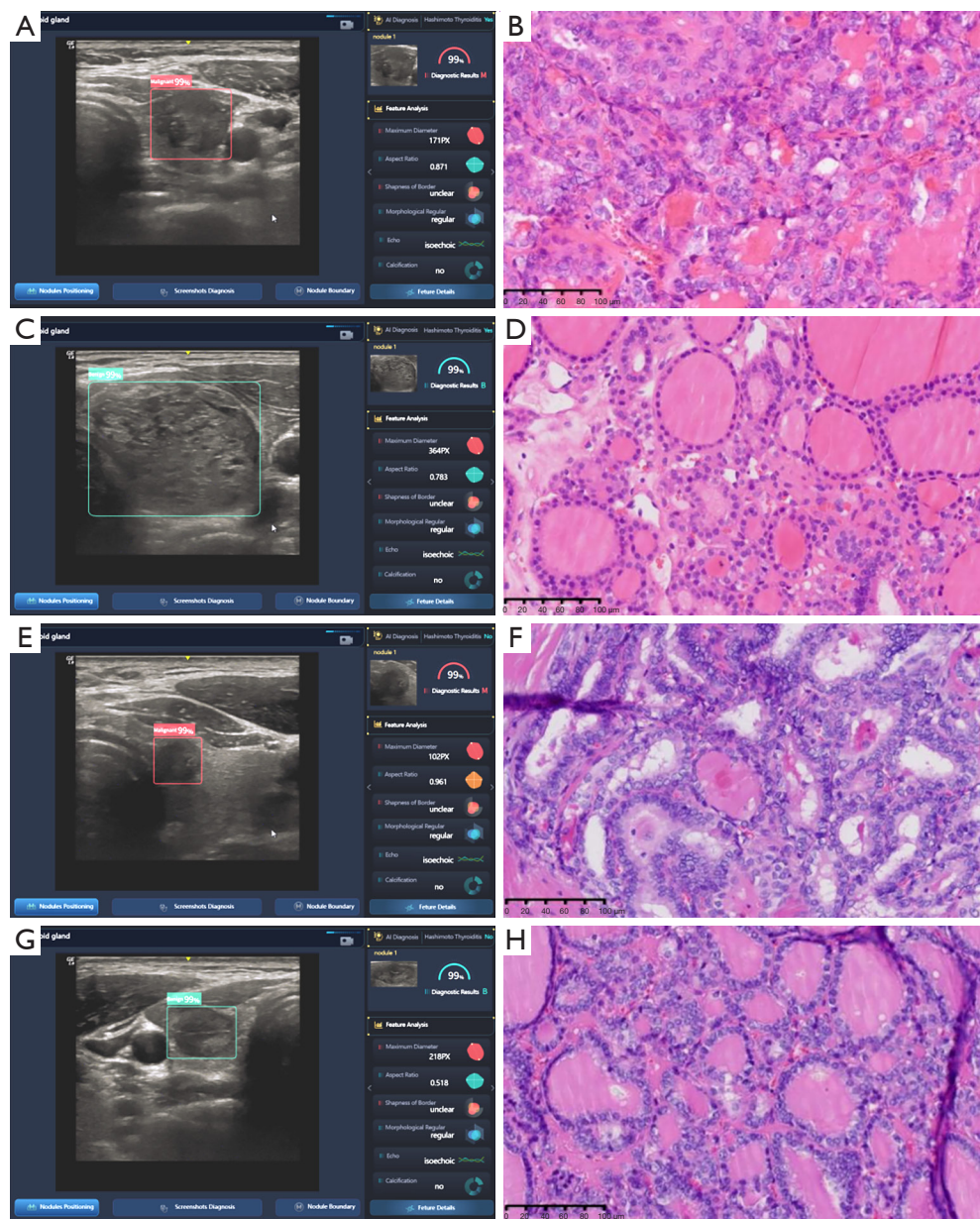


Figure 2 Typical dynamic artificial intelligence (AI) diagnostic plots and postoperative pathology. (A) Dynamic AI diagnostic plot of a malignant thyroid nodule in patients with Hashimoto thyroiditis (HT). (B) Postoperative pathology of a malignant nodule indicating thyroid papillary carcinoma [hematoxylin and eosin (HE) stain, 200 \times]. (C) Dynamic AI diagnostic plot of a benign thyroid nodule in patients with HT. (D) Postoperative pathology of a benign nodule indicating thyroid follicular adenoma (HE stain, 200 \times). (E) Dynamic AI diagnostic plot of a malignant thyroid nodule in patients without HT. (F) Postoperative pathology of a malignant nodule indicating thyroid papillary carcinoma (HE stain, 200 \times). (G) Dynamic AI diagnostic plot of a benign thyroid nodule in patients without HT. (H) Postoperative pathology of a benign nodule indicating thyroid follicular adenoma (HE stain, 200 \times).

the non-HT patients, dynamic AI diagnosed 383 malignant and 130 benign nodules, which highly corresponded to the postoperative pathological consequences [$\kappa=0.694$ (high

consistency); $P<0.001$; *Table 1*]. There existed no significant differences among the sensitivity, specificity, accuracy, PPV, NPV, missed diagnosis rate, or misdiagnosis rate between

Table 1 Consistency analysis of various detection methods and the results of pathological examinations

Detection methods	Postoperative pathology		κ	P value
	Malignant	Benign		
Dynamic AI in all patients			0.690	<0.001
Malignant	564	41		
Benign	58	166		
Dynamic AI in HT patients			0.668	<0.001
Malignant	181	16		
Benign	10	36		
Dynamic AI in non-HT patients			0.694	<0.001
Malignant	383	42		
Benign	31	130		
ACR TI-RADS in HT patients			0.569	<0.001
Malignant	186	11		
Benign	19	27		
FNAC in HT patients			0.309	<0.001
Malignant	158	15		
Benign	2	10		
Dynamic AI in HT patients with ACR TI-RADS category 4 thyroid nodules			0.528	<0.001
Malignant	41	4		
Benign	7	10		

ACR, American College of Radiology; AI, artificial intelligence; FNAC, fine needle aspiration cytology; HT, Hashimoto thyroiditis; TI-RADS, Thyroid Imaging Reporting and Data System.

Table 2 Diagnostic efficacy of dynamic artificial intelligence examination

Detection methods	Sensitivity, % (n)	Specificity, % (n)	Accuracy, % (n)	PPV, % (n)	NPV, % (n)	Missed diagnosis rate, % (n)	Misdiagnosis rate, % (n)
Dynamic AI	90.68 (564/622)	80.19 (166/207)	88.06 (730/829)	93.22 (564/605)	74.11 (166/224)	9.32 (58/622)	19.81 (41/207)

AI, artificial intelligence; PPV, positive predictive value; NPV, negative predictive value.

HT and non-HT patients (*Table 4*).

Diagnostic efficacy of dynamic AI versus ACR TI-RADS in patients with HT

Among the 243 nodules in patients with HT, ACR TI-RADS indicated 46 benign and 197 malignant nodules, of which 27 benign and 186 malignant nodules corresponded to the postoperative pathological consequences [$\kappa=0.569$ (moderate consistency); $P<0.001$; *Table 1*]. The specificity was more elevated and the rate of misdiagnosis was lower for the sake of dynamic AI compared to preoperative ACR

TI-RADS and ultrasound (*Table 5*).

Diagnostic efficacy of dynamic AI versus FNAC in patients with HT

Of the 243 nodules in patients with HT, 185 nodules underwent ultrasound-guided fine-needle biopsy before surgery. FNAC indicated 12 benign and 173 malignant nodules, among which 10 benign and 158 malignant nodules corresponded to the postoperative pathological consequences. Under dynamic AI examination, 7 benign and 168 malignant nodules corresponded to the postoperative pathological

Table 3 Stability analysis of dynamic artificial intelligence examination in patients with Hashimoto thyroiditis

Variables	Accuracy, % (n)	P value
Sex		0.302
Male	84.62 (33/39)	
Female	90.20 (184/204)	
Age (years)		0.486
<55	89.95 (179/199)	
≥55	86.36 (38/44)	
Nature of tumor		0.007
Benign	78.26 (36/46)	
Malignant	91.88 (181/197)	
Tumor size (maximum diameter)		0.262
≤0.5 cm	85.23 (75/88)	
0.5 cm < d ≤ 1.0 cm	90.20 (92/102)	
1.0 cm < d ≤ 2.0cm	97.22 (35/36)	
>2.0 cm	88.24 (15/17)	

d, diameter.

Table 4 Diagnostic efficacy of dynamic artificial intelligence examination in patients with and without Hashimoto thyroiditis

Variables	HT, % (n)	NoHT, % (n)	P value
Sensitivity	91.88 (181/197)	90.12 (383/425)	0.482
Specificity	78.26 (36/46)	80.75 (130/161)	0.709
Accuracy	89.30 (217/243)	87.54 (513/586)	0.477
PPV	94.76 (181/191)	92.51 (383/414)	0.306
NPV	69.23 (36/52)	75.58 (130/172)	0.360
Missed diagnosis rate	8.12 (16/197)	9.88 (42/425)	0.482
Misdiagnosis rate	21.74 (10/46)	19.25 (31/161)	0.709

HT, Hashimoto thyroiditis; PPV, positive predictive value; NPV, negative predictive value.

consequences [$\kappa=0.309$ (poor consistency); $P<0.001$; *Table 1*]. Compared with FNAC diagnosis, dynamic AI had a significantly higher sensitivity and lower missed diagnosis rate (*Table 6*).

Diagnostic efficacy of dynamic AI for TR 4 nodules in patients with HT

Of the 243 nodules in HT patients, 62 were TR 4 nodules. Among these, dynamic AI identified 17 benign and 45

Table 5 Diagnostic efficacy of dynamic artificial intelligence versus ACR TI-RADS in patients with Hashimoto thyroiditis

Variables	Dynamic AI, % (n)	ACR TI-RADS, % (n)	P value
Sensitivity	91.88 (181/197)	94.42 (186/197)	0.319
Specificity	78.26 (36/46)	58.70 (27/46)	0.043
Accuracy	89.30 (217/243)	87.65 (213/243)	0.570
PPV	94.76 (181/191)	90.73 (186/205)	0.124
NPV	69.23 (36/52)	71.05 (27/38)	0.852
Missed diagnosis rate	8.12 (16/197)	5.58 (11/197)	0.319
Misdiagnosis rate	21.74 (10/46)	41.30 (19/46)	0.043

ACR, American College of Radiology; TI-RADS, Thyroid Imaging Reporting and Data System; AI, artificial intelligence; HT, Hashimoto's thyroiditis; PPV, positive predictive value; NPV, negative predictive value.

Table 6 Diagnostic efficacy of dynamic artificial intelligence versus fine needle aspiration cytology in patients with Hashimoto thyroiditis

Variables	Dynamic AI, % (n)	FNAC, % (n)	P value
Sensitivity	97.10 (168/173)	91.33 (158/173)	0.021
Specificity	58.33 (7/12)	83.33 (10/12)	0.178
Accuracy	94.59 (175/185)	90.81 (168/185)	0.162
PPV	97.11 (168/173)	98.75 (158/160)	0.297
NPV	58.33 (7/12)	40.00 (10/25)	0.295
Missed diagnosis rate	2.89 (5/173)	8.67 (15/173)	0.021
Misdiagnosis rate	41.67 (5/12)	16.67 (2/12)	0.178

AI, artificial intelligence; FNAC, fine needle aspiration cytology; PPV, positive predictive value; NPV, negative predictive value.

malignant nodules, of which 10 benign and 41 malignant nodules corresponded to the postoperative pathological consequences [$\kappa=0.528$ (moderate consistency); $P<0.001$; *Table 1*]. Diagnostic efficacy of dynamic AI for TR 4 nodules in patients with HT is summarized in *Table 7*.

Discussion

Thyroid cancer is a common malignant tumor of the endocrine system, and its incidence rate is increasing year by year. In 2020, there were approximately 586,000 new

Table 7 Diagnostic efficacy of dynamic artificial intelligence for ACR TI-RADS category 4 thyroid nodules in patients with Hashimoto thyroiditis

Detection methods	Sensitivity, % (n)	Specificity, % (n)	Accuracy, % (n)	PPV, % (n)	NPV, % (n)	Missed diagnosis rate, % (n)	Misdiagnosis rate, % (n)
Dynamic AI	85.42 (41/48)	71.43 (10/14)	82.26 (51/62)	91.11 (41/45)	58.82 (10/17)	14.58 (7/48)	28.57 (4/14)

ACR, American College of Radiology; TI-RADS, Thyroid Imaging Reporting and Data System; AI, artificial intelligence; PPV, positive predictive value; NPV, negative predictive value.

cases of thyroid cancer worldwide (449,000 in women, 137,000 in men). The age-standardized incidence rate of thyroid cancer is 10.1/100,000 in women and 3.1/100,000 in men (21). HT is closely related to the occurrence of thyroid cancer (5-7). Due to the varying degrees of destruction and disappearance of follicular cells, fibrosis, and lymphocyte infiltration in the thyroid parenchyma of HT patients with nodules, the growth mode of thyroid nodules is irregular, making it difficult to judge whether thyroid nodules are benign or malignant. One study has shown that the accuracy of physicians in diagnosing thyroid cancer in patients with HT is only 54.28% (8). In addition, the diagnosis of thyroid nodules by FNAC in patients with HT is more likely to lead to unqualified samples or undetermined significance due to cytopathological changes, increasing the missed diagnosis rate and misdiagnosis rate; thus, diagnostic efficacy is low (22). Therefore, better and more effective methods are needed in the clinic to determine whether nodules are benign or malignant in patients with HT.

The joint utilization of the medical imaging and AI on the foundation of further learning has innovated the medical image interpretation. The system of AI ultrasonic intelligent auxiliary diagnosis, on the foundation of the static ultrasonic images, could quickly and conveniently automatically recognize and outline nodules and discriminate between benign and malignant nodules (13-15). Lu *et al.* (23) retrospectively studied 234 thyroid nodules that had been confirmed by surgery and pathology. In their study, the AmCAD-UT system was used to analyze preoperative ultrasound images of nodules and was found to have a diagnostic sensitivity, specificity, and accuracy of 89.0%, 82.5%, and 86.8%, respectively. Although static AI provides a new method for the thyroid nodule diagnosis, acquiring a single image is still affected by actions of individual doctors. On this basis, through dynamic AI based on deep learning, real-time localization, outlining, and diagnosis of nodules are possible because the system comprehensively characterizes the nodules and discriminates between benign and malignant nodules. In our previous study (17), we

found dynamic AI owned an elevated value in the thyroid nodule diagnosis, with a sensitivity of 92.21%, a specificity of 83.20%, an accuracy of 89.97%, and a high consistency with postoperative pathology ($\kappa=0.737$; $P<0.001$). In the present study, we used dynamic AI to diagnose benign and malignant thyroid nodules against the background of HT, and studied its diagnostic value, as well as the significance of its contribution to guiding surgical treatment strategies. As to present studies, dynamic AI was employed for the purpose of assessing the status (malignant or benign) of 829 nodules of the thyroid. The accuracy, specificity and sensitivity of dynamic AI were 88.06%, 80.19% and 90.68% respectively, and the results were highly consistent with those of postoperative pathology ($\kappa=0.690$; $P<0.001$), with high diagnostic value in terms of determining the nature of nodules. When dynamic AI was used to determine the status of nodules against a background of HT, there was no significant difference in accuracy according to age, gender, and the size of nodules, indicating that dynamic AI has good stability in the diagnosis of thyroid nodules against a background of HT.

The accuracy of dynamic AI for malignant and benign nodules was 91.88% and 71.26%, respectively. This difference is statistically significant, indicating that dynamic AI is more accurate in diagnosing malignant nodules. There are several possible explanations for this. First, dynamic AI judges benign and malignant nodules based on traditional ultrasound images through deep learning of big data. In the process of deep learning, image screening and acquisition may still have an artificial bias. Second, the set of data for benign nodules that dynamic AI can collect, acquire, and learn is smaller, and only a few patients with benign nodules who meet surgical indications undergo surgical resection and pathological diagnosis. Based on these two points, we need to increase the amount of data for deep learning in the future, especially for benign nodules, to improve the dynamic AI ultrasonic intelligent auxiliary diagnosis system. In the present study, the accuracy, specificity and sensitivity of dynamic AI in diagnosing

benign and malignant nodules against a background of HT was 91.88%, 78.26%, and 89.30%, respectively, and it had high consistency with postoperative pathological consequences ($\kappa=0.668$; $P<0.001$). There was not any significant distinction in diagnostic efficacy between HT and non-HT patients, which indicates that the diagnostic efficacy of in patients with HT can still reach levels seen non-HT patients; moreover, it demonstrates that dynamic AI can accurately identify nodules and determines their status without being affected by basic pathological changes in the gland parenchyma. When combined with HT, extensive lymphocyte infiltration, eosinophilic changes in the follicular epithelium, and the proliferation of interstitial fibrous tissue in the thyroid gland can lead to increased gland volume and a hard and brittle texture, which leads to easy bleeding of the glands during operation and increased difficulty of operation as well as a higher probability of complications such as nerve injury, parathyroid injury, and postoperative bleeding. Therefore, it is particularly important to identify the surgical indications. The use of dynamic AI in the evaluation of nodule properties in patients with HT can make a considerable contribution to individualized treatment and improved prognosis.

One study reported the accuracy of thyroid nodule diagnosis by doctors in a hospital based on ultrasound images to be approximately 70% (24), while other examining patients with HT reported an accuracy of thyroid nodule diagnosis of only 54.28% (8). Diffuse changes in HT combined with nodules will affect nodule characteristics. Many benign nodules are misdiagnosed as malignant because of unclear boundaries, irregular shape, and calcification. The results of the present study showed that the diagnostic accuracy of dynamic AI in patients with HT was 89.30%, and that its diagnostic efficacy is higher in terms of specificity, with a lower misdiagnosis rate, compared to conventional preoperative ultrasound. The distinction was significant statistically, indicating that dynamic AI is preciser in the benign nodule diagnosis than is conventional preoperative ultrasound and could thus help to reduce the misdiagnosis of benign nodules. The diagnostic efficacy of dynamic AI for malignant nodules was comparable to that of ultrasound physicians in our hospital. Diagnoses made by ultrasound physicians are influenced by personal experience, operating skills, ultrasound equipment, and the presence of thyroid disease. Studies have shown that the diagnosis of thyroid nodules is more accurate in developed regions and higher-level hospitals, and when made by senior doctors (25,26). Dynamic AI can objectively evaluate

benign and malignant nodules through deep learning. The evaluation criteria are consistent, and the diagnostic accuracy can reach levels of ultrasound physicians in high-level hospitals. Moreover, Dynamic AI is safe, noninvasive, simple, efficient, and convenient. If dynamic AI is used in areas with limited access to medical technology resources or in smaller hospitals, it can make significant contributions and improve the diagnosis and treatment of thyroid nodules.

Dynamic AI provides a new method for evaluating benign and malignant nodules against a background of HT. In the present study, we evaluated the efficacy of dynamic AI in the diagnosis of TR 4 nodules in patients with HT. The ACR guidelines state that the malignant risk of TR 4 nodules is in the range 5–20% (18). The differentiation between benign and malignant category 4 nodules has long been an issue in ultrasound examination, and the difficulty increases in the presence of HT, which further affects the formulation of individualized initial treatment plans for patients. To solve these problems, we used dynamic AI to evaluate TR 4 nodules in patients with HT as indicated by conventional preoperative ultrasound with the aim of providing a new method for the diagnosis of thyroid nodules. In the present study, the sensitivity, specificity, and accuracy of dynamic AI in diagnosing TR 4 nodules against a background of HT were 85.42%, 71.43%, and 82.26%, respectively. Diagnoses using dynamic AI can provide a strong basis for surgeons to control disease using individualized and scientific diagnosis and therapy tactics.

As for present studies, in patients with HT, dynamic AI has significantly higher sensitivity and a lower missed diagnosis rate than does FNAC diagnosis, indicating that dynamic AI is more accurate than FNAC in diagnosing malignant nodules. The consistency between FNAC diagnosis and postoperative pathological results in patients with HT was poor ($\kappa=0.309$; $P<0.001$). The reason for this is that the diagnostic efficacy of FNAC was low due to the cytopathological changes in HT; dynamic AI provides a new method for evaluating benign and malignant nodules against a background of HT. In patients with HT, the thyroid gland is fragile and has a rich blood supply, which increases the risk of bleeding after puncture. It is not easy to perform FNAC in areas with limited medical technology resources or only basic-level hospitals. In addition to being safe and non-invasive, dynamic AI can improve diagnostic efficacy while ensuring the accuracy of diagnosis, thus contributing to the alleviation of patient anxiety and the conservation of medical resources and costs.

There were some limitations in this study. The data

presented in this study were from a single center, and the duration of the study and number of patients included were limited. In the future, the number of cases will be increased and a multicenter study will be performed to better and more comprehensively evaluate dynamic AI.

Conclusions

Dynamic AI has the advantages of objectivity, accuracy, safety, and efficacy. It owns an elevated worth for malignant and benign thyroid nodules in patients with HT and for TR 4 nodules. Dynamic AI can help surgeons evaluate their patients' condition, contributing to individualized diagnosis and treatment strategies. Dynamic AI is just an emerging approach for the thyroid nodule diagnosis that is worth promoting in clinical practice. We believe that continuous research will demonstrate broader applications for dynamic AI in the medical field.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the STARD checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-22-889/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-889/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted following the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of PLA General Hospital (No. S2022-441-01). Written informed consent was obtained from all participants.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with

the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. La Vecchia C, Malvezzi M, Bosetti C, Garavello W, Bertuccio P, Levi F, Negri E. Thyroid cancer mortality and incidence: a global overview. *Int J Cancer* 2015;136:2187-95.
2. Resende de Paiva C, Grønhoj C, Feldt-Rasmussen U, von Buchwald C. Association between Hashimoto's Thyroiditis and Thyroid Cancer in 64,628 Patients. *Front Oncol* 2017;7:53.
3. Staii A, Mirocha S, Todorova-Koteva K, Glinberg S, Jaume JC. Hashimoto thyroiditis is more frequent than expected when diagnosed by cytology which uncovers a pre-clinical state. *Thyroid Res* 2010;3:11.
4. Caturegli P, De Remigis A, Chuang K, Dembele M, Iwama A, Iwama S. Hashimoto's thyroiditis: celebrating the centennial through the lens of the Johns Hopkins hospital surgical pathology records. *Thyroid* 2013;23:142-50.
5. Pan J, Ye F, Yu C, Zhu Q, Li J, Zhang Y, Tian H, Yao Y, Zhu M, Shen Y, Zhu F, Wang Y, Zhou X, Guo G, Wu Y. Papillary Thyroid Carcinoma Landscape and Its Immunological Link With Hashimoto Thyroiditis at Single-Cell Resolution. *Front Cell Dev Biol* 2021;9:758339.
6. Sur ML, Gaga R, Lazăr C, Lazea C, Aldea C, Sur D. Papillary thyroid carcinoma in children with Hashimoto's thyroiditis - a review of the literature between 2000 and 2020. *J Pediatr Endocrinol Metab* 2020;33:1511-7.
7. Subhi O, Schulten HJ, Bagatian N, Al-Dayini R, Karim S, Bakhshab S, Alotibi R, Al-Ahmadi A, Ata M, Elaimi A, Al-Muhayawi S, Mansouri M, Al-Ghamdi K, Hamour OA, Jamal A, Al-Maghrabi J, Al-Qahtani MH. Genetic relationship between Hashimoto's thyroiditis and papillary thyroid carcinoma with coexisting Hashimoto's thyroiditis. *PLoS One* 2020;15:e0234566.
8. Fang Z, Jiang TA, Jin QF, Huang JY, Chen JX. Analyses of 35 cases of papillary thyroid carcinoma under the background of Hashimoto's thyroiditis. *J Med Imaging* 2011;21:1792-94+1797.
9. Crippa S, Mazzucchelli L, Cibas ES, Ali SZ. The Bethesda System for reporting thyroid fine-needle aspiration specimens. *Am J Clin Pathol* 2010;134:343-4; author reply 345.

10. Miyagi Y, Miyake T. Potential of Artificial Intelligence for Estimating Japanese Fetal Weights. *Acta Med Okayama* 2020;74:483-93.
11. Zhou J, Du M, Chang S, Chen Z. Artificial intelligence in echocardiography: detection, functional evaluation, and disease diagnosis. *Cardiovasc Ultrasound* 2021;19:29.
12. Gao Y, Liu B, Zhu Y, Chen L, Tan M, Xiao X, Yu G, Guo Y. Detection and recognition of ultrasound breast nodules based on semi-supervised deep learning: a powerful alternative strategy. *Quant Imaging Med Surg* 2021;11:2265-78.
13. Nguyen DT, Pham TD, Batchuluun G, Yoon HS, Park KR. Artificial Intelligence-Based Thyroid Nodule Classification Using Information from Spatial and Frequency Domains. *J Clin Med* 2019.
14. Han M, Ha EJ, Park JH. Computer-Aided Diagnostic System for Thyroid Nodules on Ultrasonography: Diagnostic Performance Based on the Thyroid Imaging Reporting and Data System Classification and Dichotomous Outcomes. *AJNR Am J Neuroradiol* 2021;42:559-65.
15. Zhu J, Zhang S, Yu R, Liu Z, Gao H, Yue B, Liu X, Zheng X, Gao M, Wei X. An efficient deep convolutional neural network model for visual localization and automatic diagnosis of thyroid nodules on ultrasound images. *Quant Imaging Med Surg* 2021;11:1368-80.
16. Zong ST, Hu MX, Xu N, Li JZ, Xia CY, Yu FX, Zhu Q, Zhao HX. The application of thyroid CAD system in the diagnosis of thyroid nodules concurrent with Hashimoto's thyroiditis: preliminary study. *Chin Arch Otolaryngol Head Neck Surg* 2020;27:512-5.
17. Wang B, Wan Z, Li C, Zhang M, Shi Y, Miao X, Jian Y, Luo Y, Yao J, Tian W. Identification of benign and malignant thyroid nodules based on dynamic AI ultrasound intelligent auxiliary diagnosis system. *Front Endocrinol (Lausanne)* 2022;13:1018321.
18. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, Hammers LW, Hamper UM, Langer JE, Reading CC, Scoutt LM, Stavros AT. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol* 2017;14:587-95.
19. Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid* 2017;27:1341-6.
20. Baloch ZW, Asa SL, Barletta JA, Ghossein RA, Juhlin CC, Jung CK, LiVolsi VA, Papotti MG, Sobrinho-Simões M, Tallini G, Mete O. Overview of the 2022 WHO Classification of Thyroid Neoplasms. *Endocr Pathol* 2022;33:27-63.
21. Pizzato M, Li M, Vignat J, Laversanne M, Singh D, La Vecchia C, Vaccarella S. The epidemiological landscape of thyroid cancer worldwide: GLOBOCAN estimates for incidence and mortality rates in 2020. *Lancet Diabetes Endocrinol* 2022;10:264-72.
22. Meng C, Hinkle LE, Wang W, Su D, Li X. Hashimoto's thyroiditis elicits decreased diagnostic efficacy of thyroid nodule ultrasound-guided fine needle aspiration. *Int J Clin Exp Pathol* 2019;12:3474-82.
23. Lu Y, Shi XQ, Zhao X, Song D, Li J. Value of Computer Software for Assisting Sonographers in the Diagnosis of Thyroid Imaging Reporting and Data System Grade 3 and 4 Thyroid Space-Occupying Lesions. *J Ultrasound Med* 2019;38:3291-300.
24. Kim DW, Lee EJ, In HS, Kim SJ. Sonographic differentiation of partially cystic thyroid nodules: a prospective study. *AJNR Am J Neuroradiol* 2010;31:1961-6.
25. Jin Z, Zhu Y, Zhang S, Xie F, Zhang M, Zhang Y, Tian X, Zhang J, Luo Y, Cao J. Ultrasound Computer-Aided Diagnosis (CAD) Based on the Thyroid Imaging Reporting and Data System (TI-RADS) to Distinguish Benign from Malignant Thyroid Nodules and the Diagnostic Performance of Radiologists with Different Diagnostic Experience. *Med Sci Monit* 2020;26:e918452.
26. Ji Q, Gao YT, Zhou YF, Jing HB, Lu CY, Wang JF, He D, Chen GF, Xu SH, Wu XP, Liu C. A comparison of diagnosis and treatment status of thyroid nodules in two hospitals with different classes. *Jiangsu Med J* 2020;46:1008-11.

Cite this article as: Wang B, Wan Z, Zhang M, Gong F, Zhang L, Luo Y, Yao J, Li C, Tian W. Diagnostic value of a dynamic artificial intelligence ultrasonic intelligent auxiliary diagnosis system for benign and malignant thyroid nodules in patients with Hashimoto thyroiditis. *Quant Imaging Med Surg* 2023;13(6):3618-3629. doi: 10.21037/qims-22-889